神经模拟推断：基于神经网络和模拟推断的认知建模方法

潘晚坷[1,2]，胡传鹏[1,2]

[1]南京师范大学心理学院，南京 210097

[2]江苏省高校哲学社会科学实验室——南京师范大学青少年教育与智能支持实验室

摘要

认知建模是量化和理解人类心智过程的重要方法，但目前该方法多集中于简单实验任务和数据结构，当试图构建复杂模型以解释复杂认知过程时，能否确定其似然并完成参数推断成为一项严峻的挑战。基于神经网络的模拟推断方法结合了模拟推断与分摊技术，无需计算似然函数，直接利用模拟数据来参数推断，并通过神经网络训练控制计算成本，可快速稳健地进行参数推断。该方法已成功应用于证据积累模型框架下的大规模数据、动态潜变量以及联合建模等场景，并开始扩展到强化学习和贝叶斯决策模型。未来的研究可以进一步验证神经模拟推断的有效性，应用该方法扩大认知模型的应用范围，推动理论与模型发展，增进对人类复杂认知加工规律的理解。

关键词 认知建模，生成模型，贝叶斯，基于模拟的推断，基于神经网络的模拟推断

1. 引言

认知计算建模通过构建形式化数学模型来量化人类心智过程(Churchland & Sejnowski, 1988; McClelland, 2009)。该方法通过模型参数将抽象的认知机制转化为可验证的计算假设，能够整合多模态数据、量化个体差异，并为精神病理学的计算诊断提供依据(Eaton et al., 2023; Geng et al., 2022; Kriegeskorte & Douglas, 2018; Wang et al., 2014)。

现代认知建模遵循“生成-推断-验证”的迭代研究范式(Guest & Martin, 2021)，大致包含四个相互关联的环节（见图 1）：(1)模型定义，将符号化理论模型形式化为生成式计算模型（见附录 1.1）；(2)数据收集，根据计算模型设计实验并收集数据；(3)参数推断，根据观测数据推断模型参数，从而量化认知过程；(4)模型评估与比较，通过后验预测检验、参数恢复和模型恢复等方式验证模型的表现。如果有多个潜在的模型，则通过模型比较来检验不同的理论假设（见郭鸣谦等, 2024；附录 1.2）。关于认知建模更详细的步骤，请参考(Guest & Martin, 2021; Schad et al., 2021; Wilson & Collins, 2019)。参数推断是连接模型与观测数据的核心。在实践中，它不仅是对认知机制进行量化的必要步骤，还为模型验证和比较提供了基础。在一定程度上，参数推断也可能会对实验设计与数据收集产生制约：当无法对复杂的模型进行参数推断时，研究者只能选择简化模型，并据此选择简化的实验设计和解释理论。

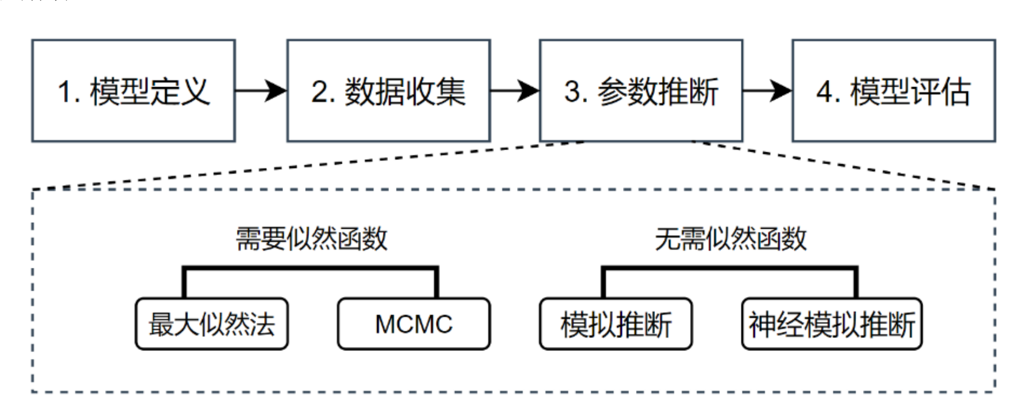


图 1 **认知建模流程与参数推断方法的示意图**。认知建模包括四个基本流程，其中参数推断是连接模型与观测数据的核心。参数推断可根据对于似然函数的依赖性分为两类：一类是需要似然函数的方法，包括最大似然法和 MCMC；另一类是无需似然函数的方法，涵盖早期的模拟推断以及近期提出的神经模拟推断。

传统参数推断方法可分为两大体系：频率学派的最大似然估计法(Maximum Likelihood Estimation, MLE)及其衍生的相关方法；贝叶斯学派的马尔可夫链蒙特卡洛(Markov Chain Monte Carlo, MCMC)和变分推断等。尽管这些方法在不同场景下具有各自的应用价值，但其均依赖显式似然函数(explicit likelihood function)的数学可处理性(analytical tractability)，且存在各自的局限性（见表 1）。当面对复杂场景时，似然函数可能面临无法获取或计算成本高昂的挑战。具体而言，当模型包含复杂动态机制时，例如在学习反馈情景中，对刺

激的加工表现会在不同试次间出现非线性变化(Miletić et al., 2021)，此时解析似然函数很难推导得出；当数据结构呈现多层级特性，例如在处理多个被试的分层强化学习模型(van Geen & Gerraty, 2021)时，最大似然估计难以应对同时包含就个体和群体参数的分层模型，而 MCMC 估计在处理数据量过大时存在速度过慢的问题，其速度与数量呈线性增长(Ahn et al., 2017)；当对多个模态的数据进行联合建模，如实现试次级别的 EEG 信号与认知模型的联合建模(Nunez et al., 2019)时，额外变量将导致似然函数的计算成本高昂。

表 1 不同参数推断方法对比

	最大似然估计	MCMC	早期模拟推断	神经模拟推断
似然需求	显式解析式	显式可计算	无需显式形式	无需显式形式
计算效率	高（单次优化）	低（迭代采样）	中等（模拟驱动）	双阶段分化（预训练阶段低，推理极端高）
输出形式	点估计	后验分布	近似后验分布	近似后验分布
维度诅咒	严重（梯度消失）	严重（混合速度下降）	中等（模拟次数限制）	部分缓解（受限于预训练数据覆盖度）
典型应用场景	标准强化学习模型、漂移扩散模型	分层强化学习模型	具有时变漂移率的漂移扩散模型	时变漂移率的漂移扩散模型
工具包	Python-scipy (Virtanen et al., 2020) R-optim (Varadhan, 2014) Matlab-fminunc (Ponce-Ortega et al., 2024)	PyMC (Abril-Pla et al., 2023) Stan (Gelman et al., 2015)	ABCtoolbox (Wegmann et al., 2010)	Bayesflow (Radev et al., 2023) SBI (Tejero-Cantero et al., 2020)

注：似然函数的依赖形式有三类(Bürkner et al., 2023)。其中，显式解析式指似然函数具备明确数学表达式，适用于解析推导与高效优化；显式可计算虽缺乏简洁公式，但可通过数值方法计算具体值；无需显式形式则完全摆脱对具体公式的依赖，借助模拟采样或神经网络学习分布特性。显式程度越高，模型假设越严格但计算越可控；而隐式方法虽增加复杂度，却显著扩展了传统模型在高维及复杂动态系统中的应用范围(Bürkner et al., 2023)。

似然函数计算的挑战催生了基于模拟的推断(Simulation-based Inference, SBI)，简称为模拟推断，其核心思想在于通过生成模拟数据集，构建参数与数据的随机映射关系，从而实现逆向推断(Beaumont, 2010; Cranmer et al., 2020)。不同于最大似然法和 MCMC，模拟推断可以绕开对似然函数的计算。模拟推断的标准流程包含三个阶段（见附录 2）：基于参数先验分布和生成模型，生成参数和模拟数据的数据对；计算模拟数据和实际数据的差异，根据预设的阈值决定是否接受参数作为后验样本；与 MCMC 方法类似的是，模拟推断需要重复上述采样过程，以累积足够的样本来构建后验分布的近似(Palestro et al., 2018)。

虽然早期的模拟推断方法能够解决似然计算的挑战，但其每次后验参数采样时均需生成模拟数据，不可避免地产生大量的计算冗余(Radev et al., 2021)。此外，该方法还受到数据和参数维度的限制，在处理大规模数据集和高度复杂的模型时面临着计算效率低下的问题。

为解决模拟推断的这些问题，研究者将神经网络和深度学习方法引入模拟推断(Cranmer et al., 2020)，本文将这类方法称为“基于神经网络的模拟推断”(Neural Simulation-based Inference, NSBI)，简称为“神经模拟推断”。神经模拟推断通过引入分摊推理(amortized inference)机制（见图 2），显著提升了认知建模的推断效率(Radev et al., 2021)。在早期模拟推断方法的基础上，神经模拟推断通过海量模拟数据训练神经网络学习参数与数据间的映射关系，从而达到在推断阶段对新观测数据执行快速后验估计(Radev et al., 2021)。神经模拟推断在预训练阶段需投入较多时间，但一旦训练完毕，便能实现迅速推理。例如，仅需一天的训练，神经模拟推断便能在大约 5 秒内完成 500 个数据集的推理任务，而早期模拟推断方法可能需要数百个小时(Radev et al., 2022)。随着工具的开发，神经模拟推断的应用门槛已显著降低，使得该方法与传统推断方法的易用性趋近，例如 BayesFlow 和 SBI（表 1）。两者均基于成熟的深度学习框架，BayesFlow 依托 TensorFlow，SBI 基于 PyTorch，通过简洁的 API 设计简化了概率编程流程，但要求用户具备基础的神经网络训练与概率建模知识(Cranmer et al., 2020)。在计算资源方面，神经网络预训练阶段需中等性能 GPU 支持（如 NVIDIA 1080Ti），训练数据的模拟则依赖 CPU 算力，后者更可能成为效率瓶颈(Radev et al., 2022)。尽管功能覆盖相似，两者存在侧重差异，BayesFlow 在认知模型和流行病学模型的使用更多， SBI 则 在生物医学和生态学模型中应用更广。

神经模拟推断已经获得更多的关注和应用(Boelts et al., 2022; Fengler et al., 2021; Ghaderi-Kangavari et al., 2023; Schumacher et al., 2024; von Krause et al., 2022)，但其技术原理晦涩，应用场景多样，缺乏系统化的梳理和实践指导。本文首先详细阐述分摊推理的数学基础及其在神经网络中的实现机制，随后通过具体应用案例探讨神经模拟推断的实际意义，最后分析其局限性及未来的发展潜力。

2. 神经模拟推断的技术革新

神经模拟推断已发展出三个主要方向（见附录 4）：神经后验估计，能够直接从输入数据中输出参数的后验分布；神经似然估计，根据数据和参数输出模型对应的似然值；以及基于神经网络的模型比较，能够基于输入数据评估多个认知模型产生该数据的概率或贝

叶斯因子。鉴于神经后验估计在应用上的广泛性，本研究将重点阐述该方法的基本原理及其在实际中的应用。

神经后验估计专注于直接对参数的后验分布进行估计，其通过直接学习参数的后验分布，避免了传统方法如 MCMC 等间接推断的复杂性和计算成本。在神经网络训练阶段，首先需要参数空间分布中采样参数$\tilde{\theta}_i \sim \pi(\theta)$，然后根据生成模型模拟数据$\widetilde{D}_i^{1:M} \sim p(D|\theta)$，从而可以生成大量训练数据对$\{\tilde{\theta}_i, \widetilde{D}_i^{1:M}\}_{i=1}^{N}$，其中$i$为参数样本索引（共$N$个参数样本），$M$代表每个参数样本对应模拟生成的数据数量。随后，神经网络将$\widetilde{D}_i^{1:M}$作为输入，输出对于参数的预测θ_i，通过调整神经网络权重ϕ最小化预测参数θ_i和训练参数$\tilde{\theta}_i$的差距，使得神经网络$q_\phi(\theta|D)$逼近真实的参数后验概率密度$p(\theta|D)$。在推理阶段，将真实观测数据D输入到训练好的神经网络，可以根据神经网络所估计的后验分布进行参数采样$\theta_i \sim q_\phi(\theta|D)$。这种方法不仅简化了推断过程，还提高了计算效率，适用于高维参数空间和复杂模型结构的场景。

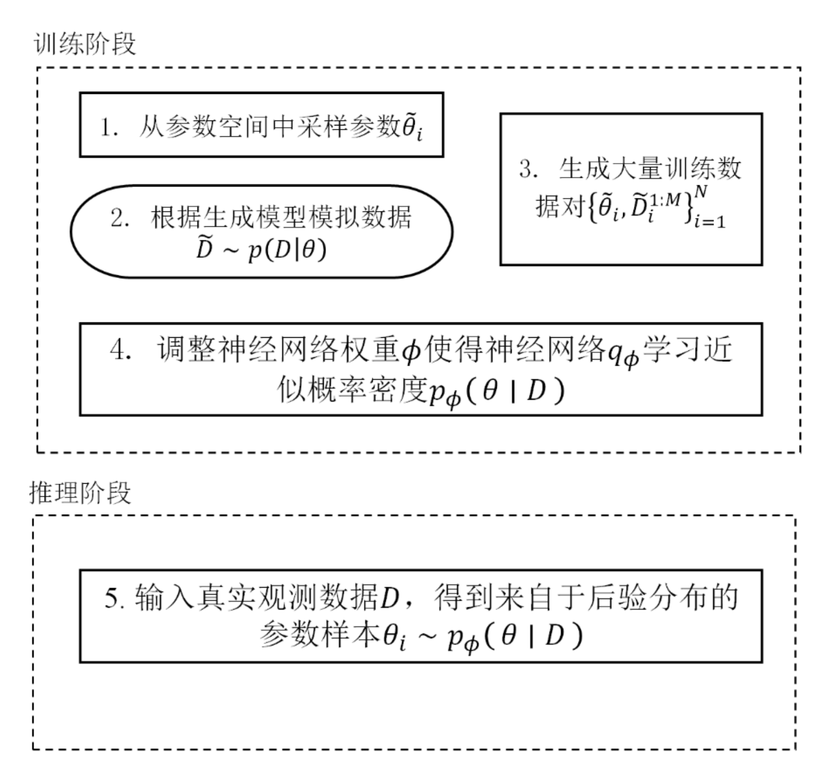


图 2 **基于神经网络的模拟推断方法算法流程**。椭圆形图示代表的是模拟数据生成的过程，这一环节通常消耗最多的计算资源。与早期模拟方法相比，采用基于神经网络的模拟推理方法能在训练阶段分散大量模拟数据的计算需求，因此在推理阶段无需再生成模拟数据即可进行推理，显著提高了效率。早期模拟推断方法的具体计算过程请参考附录 2。

Radev 等首次将神经后验估计方法应用于认知建模领域，提出了 BayesFlow 的框架 (Radev et al., 2022)。BayesFlow 采用基于标准化流（详见附录 3.2）的神经网络架构。在这种架构中，认知模型的参数后验分布 $p(\theta \mid D)$可以表示为在给定数据D的条件下参数θ和潜在变量z_θ（来自于简单分布）之间的标准化流$p_\phi\left(\theta \,\middle|\, \mathcal{H}_\eta(D)\right) = p(z_\theta)\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right|$。此

外，BayesFlow 引入摘要统计网络(summary statistics network)，即$\mathcal{H}_\eta$，以解决先前神经后验估计中只能对特定试次数量数据进行参数推断的问题。通过摘要统计网络，任意长度的数据$D^{1:M}$（假设这些数据来自于独立同分布，并且具有置换不变性）可以被映射为固定长度的嵌入向量$\mathcal{H}_\eta(D)^{1:K}$，其中$K$代表摘要统计维度的潜在维度。

BayesFlow 采用无监督学习方式（详见附录 3.5），能够直接从输入数据中学习关于参数的条件概率分布，从而实现对真实后验分布的有效近似(Radev et al., 2022)。BayesFlow 的最大优势在于它能够直接对参数进行估计，而不需要对数据大小和后验分布的形态做出预先假设。最后， BayesFlow 能够实现高度并行的计算。依据 BayesFlow 框架，已经开发对应的 Python 软件包，得益于现代 GPU 的加速，研究者可快速实施和应用神经后验估计方法(Radev, Schmitt, Schumacher, et al., 2023)。神经模拟推断的框架和工具已取得持续进展，更多详细信息请参阅(Zammit-Mangion et al., 2024)。

3.   神经模拟推断在心理学研究中的应用

神经模拟推断框架和工具的持续发展，使其易用性迅速提升，尤其是神经后验估计技术方面。本文将以证据积累模型为例，介绍神经后验估计在大规模数据处理、复杂认知模型构建以及与多模态数据联合建模等场景下的扩展模型应用。通过这些应用实例，本文将阐述其优势，包括参数推断的高效性、允许更灵活的模型设定，以及对模型进行更全面检验与验证的能力。

3.1.   大规模认知老化研究数据应用

大规模数据集能够揭示在小样本研究中难以捕捉的细微模式和变异(Hedge et al., 2018; Kucina et al., 2023)，对于检验认知模型的有效性具有重要意义。但使用传统参数推断方法耗时过长，难以满足实际需求，还可能因为极端数据的偏差影响模型的可靠性。神经模拟推断通过分摊方式带来的高效性则能有效解决认知模型在大规模数据集中应用的问题。

von Krause 等人(2022)对来自 120 万参与者的在线内隐联想测验数据进行深入分析，旨在探讨老龄化对认知能力的影响。鉴于数据量庞大，传统参数统计方法难以应对，该研究者采用神经模拟推断来解决这一难题。该研究以漂移扩散模型为基础，设定模型中漂移率和决策阈值会随着内隐联想测验的实验条件（一致 vs 不一致）而变化，并区分了非决策时间参数在正确和错误试次下的变化，从而能更准确地分离出实际决策阶段的认知过程。此外，数据通过线上收集，虽然被试数量众多，但单个被试仅有 60 次反应，不足以有效估计模型的试次间变异性参数(Boag et al., 2024)。因此，为降低模型复杂度，将该参数固定为 0。

在参数推断时，von Krause 等人(2022)利用神经后验估计技术的分摊特性加速参数推断。在训练阶段，利用 GPU 加速的笔记本电脑，仅耗时约 8 小时即可完成训练；而在推理阶段，在没有 GPU 加速的机器上对整个数据集进行推理，所需时间约为 24 小时。研究者将神经模拟推断与传统的 MCMC 方法进行了对比，通过对随机抽取的 6400 名被试数据进行分析（耗时大约 53 个小时），结果显示，基于 MCMC 和神经模拟推断的参数相关系数一致性均在 0.9 以上，充分验证了该方法的稳健性。

von Krause 等人(2022)通过对后验分布的细致分析，发现认知速度随年龄增长呈现的非线性变化趋势：在 30 岁左右达到峰值，随后在 60 岁左右开始下降，这一发现与传统的基于平均反应时的研究结果形成了显著对比。这一研究表明，神经模拟推断扩大了认知建模的范围，使得对大规模数据集的认知建模成为可能。

3.2. 动态决策模型中的应用

证据积累模型（如漂移扩散模型）通常预设信息积累速率和决策边界等参数在实验过程中保持恒定(Ratcliff et al., 2008)。这种预设旨在简化模型，以便顺利完成参数推断的过程。然而，最近的研究表明，认知加工在实验过程中并非恒定，而是处于动态变化之中(见刘逸康, 胡传鹏, 2024)。例如，决策阈值可能因疲劳而逐渐降低(Voskuilen et al., 2016)，或信息积累速度可能因学习而持续提升(Miletić et al., 2021; Pedersen et al., 2017)。如果将这些认知过程的动态特征引入到认知模型，核心在于更复杂的认知模型能够完成参数的推断。

Schumacher 等人(2023)在两项研究中，通过神经模拟推断来解决动态认知模型中参数推断的难题。在认知模型上，他们构建了神经超统计(neural superstatistics)模型架构，将认知过程分解为观测层（行为数据生成机制）与动态层（参数时变过程）(Schumacher et al., 2023, 2024)。在高水平的动态层中，模型参数在试次间进行变化，变化规律受到高水平模型和参数的影响。例如，高水平模型可以是高斯过程模型，其中包含自由核参数（如振幅）。低水平模型的参数变化则遵循由这些自由核参数定义的高斯过程。在观测层面上，假设试次水平的数据是由低水平模型生成的，而低水平模型的参数在不同试次间发生变化，从而解释了数据在试次间的变化。参数推断的过程需要同时估计高层次模型参数和试次水平的模型参数。由于涉及多层级参数的复杂交互，早期方法难以有效解决此类问题。Schumacher 等人使用神经后验估计技术，使用分层的循环神经网络结构处理超统计模型中的分层和数据依赖特性，直接构建从行为数据到参数的映射。

Schumacher 等人发现，神经模拟推断可以高效地对复杂动态认知模型进行参数推断，其参数恢复的相关性均达到 0.9 以上(Schumacher et al., 2023)。在连续决策任务中，动态模型对选择行为的预测准确率显著高于传统模型，说明动态模型更优。这些发现表明，能够捕捉认知过程动态变化的模型更贴近人类决策的实际机制，从而深化了对人类决策过程的

理解。这类模型不仅能够解释微观层面认知动态与神经信号波动之间的关系(Brosnan et al., 2020; O’Connell & Kelly, 2021)，还能揭示宏观层面生命周期中认知能力的变化规律(Schumacher et al., 2023)。神经后验估计技术在该研究中起到关键作用，表明模型复杂性可能不再是认知建模应用的瓶颈。

3.3. 神经认知联合建模的应用

认知建模中另一个常见的拓展是联合神经数据和行为数据进行建模，以揭示认知过程的神经基础。然而，先前联合建模受到参数推断的限制，仅能对组水平神经数据和行为数据的均值进行建模，进行单试次水平的联合建模难以实现。单试次水平的神经活动（如 ERP 的 N200 潜伏期、CPP 斜率）更直接反映认知过程的动态，对于提升模型的解释力具有明显的价值(Ghaderi-Kangavari et al., 2023; O’Connell & Kelly, 2021)。然而，多模态数据的整合为认知模型的参数推断带来了挑战，早期的尝试发现参数推断的准确性和效率较差(Ghaderi-Kangavari et al., 2023)。神经模拟推断通过结合神经网络对高维数据的强大处理能力，为此类复杂模型的参数推断提供了有效的新解决方案。

Ghaderi-Kangavari 等人(2023)提出了一种整合神经认知建模框架，旨在联合描述单试次 EEG 与决策行为的关系。具体而言，该框架下定义的认知模型包含动态潜变量参数以及根据潜变量生成神经和行为数据的生成机制。例如，模型假设 CPP 斜率与证据累积速度相关，进而在生成模型中，证据累积速率参数漂移率会同时影响行为数据和 CPP 斜率。反过来，行为数据和 CPP 斜率在每个试次上的差异也导致了漂移率在试次间的变化。利用神经后验估计，研究者可以仅输入神经和行为数据，而直接获得潜变量参数，从而避开了联合建模中似然的不可解析性所带来的障碍。

研究结果证实了模型的准确性和稳健性，即使在数据违反正态假设或存在污染试次的情况下，参数估计仍保持稳定。在真实数据应用中，模型成功解释了非决策时间的变异，并揭示了 CPP 斜率与漂移率之间的关联，表明 CPP 可作为证据积累的神经标记(Ghaderi-Kangavari et al., 2023)。与已有模型(Nunez et al., 2017)相比，单试次整合模型在参数解释力和预测能力上更胜一筹。该研究进一步表明，神经模拟推断通过将参数推断转化为神经网络优化问题，能有效克服数据复杂性对模型定义和参数推断的制约，为对认知建模中更复杂的多模态数据分析提供基础。

4. 总结和展望

认知建模作为理解复杂心理过程的工具，正日益获得重视(Eaton et al., 2023; Geng et al., 2022; Kriegeskorte & Douglas, 2018; Wang et al., 2014)。然而，参数推断的局限性制约了认知模型更广泛的应用，实践上也限制了更复杂的实验设计及其能解决的理论争议。基于神

经网络的模拟方法为突破这一瓶颈提供了新的解决方案(Zammit-Mangion et al., 2024)。其核心优势在于：允许更灵活的模型开发，从而避免复杂的数学公式推断；高效处理高维和复杂结构的数据；利用分摊特性实现高效推理。然而，要在认知建模中广泛采用基于神经网络的模拟推断，仍面临一些挑战。

4.1. 挑战

神经模拟推断方法的有效性依赖于生成认知模型的有效性，不准确的生成模型设定会对神经网络的推断性能造成严重影响(Cannon et al., 2022)。例如，生成模型中对数据生成过程的错误假设可能导致参数或者模型无法识别。若生成模型未能准确捕捉到数据生成过程中的关键因素，如忽略某些重要的变量或错误地设定变量之间的关系，神经网络在训练过程中将无法正确学习到数据的内在结构，从而导致参数估计偏差或模型失效(Frazier et al., 2020)。目前，尚不清楚神经模拟推断方法是否能提高参数的恢复性，也不明确参数或模型的不可识别性对神经模拟推断的影响。

当前，神经模拟推断算法还可能存在过度拟合的问题，进而导致生成过于自信的后验估计结果(Hermans et al., 2022)。尽管神经网络在处理训练样本时能够提供稳健的估计，但当面对新数据或异常数据时，过度拟合的模型可能会表现出较差的泛化能力，无法有效应对数据的变化。此外，若生成模型所假设的数据生成机制与真实数据的生成机制存在差异，过拟合的神经网络将更易产生误导性估计。近期研究表明，在神经模拟推断框架下的标准漂移扩散模型中，即便数据中混入一定量的干扰数据，模拟推断方法依然展现出较强的稳健性(Wu et al., 2024)。然而，对于更复杂的模型是否具备相似特性，以及数据真实生成机制与认知模型生成机制之间的差异达到何种程度会引发神经网络产生偏差估计，仍有待未来研究进一步探索。

神经模拟推断另一个争议在于训练成本的权衡。一些研究者认为分摊策略实际上是一种对于模拟数据的浪费(Greenberg et al., 2019; Papamakarios & Murray, 2016)，因其在训练阶段需要模拟大量的的数据进行神经网络的训练。与顺序神经密度估计相比，分摊神经网络通常需要更多的模拟来充分探索参数空间。但分摊方法的优势在于一旦训练完成，即可应用于大量数据。Radev 等(2022)提出了模拟预算收支平衡的概念，用以量化在处理多大规模的数据时，非分摊方法的推理时间会超过分摊方法的训练时间。例如，在完成包含超过 75 份数据集的推理任务时，分摊方法展现出相对于其他方法的优势(Radev et al., 2022)。在处理包含数千人的大型数据集时，例如应用于心理测量模型或辅助医疗诊断模型，分摊方法具有显著的优势。此外，在比较多个结构相似的认知模型时，需要分别将每个模型重复拟合多个数据集，在此情况下，采用分摊方法更为有效。

4.2. 展望

未来的研究需要进一步开发对神经模拟推断进行有效性评估的方法，为研究者应用神经模拟推断提供方法保障。近期，研究者提出了最大平均差异(Maximum Mean Discrepancy, MMD)方法，用于识别模型错误(Schumacher et al., 2024)。该方法能够检测参数范围对数据预测与真实数据之间差异的影响，并可用于优化神经网络的训练过程。另一种提升估计稳健性的思路在于结合多种神经模拟推断方法。Radev 等(2023)提出了名为 JANA (Jointly Amortized Neural Approximation)的整合框架，旨通过神经后验网络进行参数估计，并利用参数样本结合神经似然网络计算模型比较指标。研究结果显示，JANA 相较于单一估计方法，能够获得更优的参数估计和边际似然值估计效果(Radev, Schmitt, Pratz, et al., 2023)。此外，还可进一步探索多种方法以更全面地衡量神经网络估计的性能和生成式认知模型定义的有效性，开发新型方法以提升神经网络的估计表现并增强认知模型的解释性。

由于神经模拟推断方法能克服数据复杂性和模型复杂性的双重限制，研究者将能更加灵活地开发计算模型，避免繁琐的数学推导过程，从而将认知建模应用于更多的场景之中。首先，神经模拟推断有助于开发更多包含动态机制的认知模型。已有研究已将其成功应用于证据累积模型框架下噪声为非均匀分布的 Lévy flight 模型(Hato et al., 2024)，未来研究可进一步将其扩展至其他证据积累模型的变式，如决策边界崩塌 (collapsing boundary)和泄漏竞争(leaky competing)的扩散决策模型(Tsetsos et al., 2012)。其次，数据的规模和复杂性对认知模型的限制变小。已有研究通过神经模拟推断将认知建模应用于大规模数据、神经信号与行为数据的建模（详见第 3 小节）及复杂的分层认知模型(Habermann et al., 2024)。未来研究可以进一步探索将眼动数据(Krajbich, 2019)、多种神经指标(Brosnan et al., 2020)和信心评分(Glickman et al., 2022)等数据纳入模型。最后，未研究需要将神经模拟推断拓展至更多的模型框架。例如，已有研究者使用神经模拟推断应用于强化学习模型(Rmus et al., 2024)、多项式决策树模型(Elsemüller et al., 2023)和贝叶斯决策模型(Straub et al., 2025)。但这些模型框架下的变式繁多，因此仍有诸多潜在的模型可以应用神经模拟推断方法。通过对其进行评估、识别其不足，并据此进行优化，进而推动认知建模的发展。

认知建模已经突破认知心理学的范围，开始广泛地应用于心理学及相关学科，表明其在跨领域应用的广阔潜力。神经模拟推断可能成为认知建模跨领域应用的重要推动力。例如，在计算精神病学领域，认知建模需要与更具有生态效度的实验任务和/或神经数据相结合，但传统方法无法完成这些情景下的参数推断(Ging-Jehli et al., 2021)。神经模拟推断的加入可以帮助研究者将认知模型应用于这些情景，从而加深对精神疾病潜在的认知加工异常的理解。又如，在发展与教育领域，认知模型可以帮助揭示学习能力的认知发展机制，

但通常需要构建能够同时解释反应和反应时的混合认知模型(Embretson, 2023)，也可以使神经模拟推断进行参数推断，从而让认知建模真正地应用于发展与教育领域。

参考文献

刘逸康, 胡传鹏. (2024). 证据积累模型的行为与认知神经证据. *科学通报*, *69*(8), 1068–1081. https://doi.org/10.1360/TB-2023-1080.

郭鸣谦, 潘晚坷, 胡传鹏. (2024). 认知建模中模型比较的方法. *心理科学进展*, *32*(10), 1736. https://doi.org/10.3724/SP.J.1042.2024.01736.

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry*, *1*(0), 24–24. https://doi.org/10.1162/cpsy_a_00002

Beaumont, M. A. (2010). Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 379–406. https://doi.org/10.1146/annurev-ecolsys-102209-144621

Bhattacharya, A., Sarkar, B., & Mukherjee, S. K. (2007). Distance-based consensus method for ABC analysis. *International Journal of Production Research*, *45*(15), 3405–3420. https://doi.org/10.1080/00207540600847145

Boag, R. J., Innes, R., Stevenson, N., Bahg, G., Busemeyer, J. R., Cox, G. E., Donkin, C., Frank, M., Hawkins, G., Heathcote, A., Hedge, C., Lerche, V., Lilburn, S., Logan, G. D., Matzke, D., Miletic, S., Osth, A. F., Palmeri, T., Sederberg, P. B., … Forstmann, B. (2024). *An expert guide to planning experimental tasks for evidence accumulation modelling*. OSF. https://doi.org/10.31234/osf.io/snqgp

Boelts, J., Lueckmann, J.-M., Gao, R., & Macke, J. H. (2022). Flexible and efficient simulation-based inference for models of decision-making. *eLife*, *11*, e77220. https://doi.org/10.7554/eLife.77220

Brosnan, M. B., Sabaroedin, K., Silk, T., Genc, S., Newman, D. P., Loughnane, G. M., Fornito, A., O’Connell, R. G., & Bellgrove, M. A. (2020). Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. *Nature Human Behaviour*, *4*(8), Article 8. https://doi.org/10.1038/s41562-020-0863-4

Bürkner, P.-C., Scholz, M., & Radev, S. T. (2023). Some models are useful, but how do we know which ones? Towards a unified Bayesian model taxonomy. *Statistics Surveys*, *17*(none). https://doi.org/10.1214/23-SS145

Cannon, P., Ward, D., & Schmon, S. M. (2022). *Investigating the Impact of Model Misspecification in Neural Simulation-based Inference* (No. arXiv:2209.01845). arXiv. http://arxiv.org/abs/2209.01845

Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, *242*(4879), 741–745. https://doi.org/10.1126/science.3055294

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences, 117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Eaton, N. R., Bringmann, L. F., Elmer, T., Fried, E. I., Forbes, M. K., Greene, A. L., Krueger, R. F., Kotov, R., McGorry, P. D., Mei, C., & Waszczuk, M. A. (2023). A review of approaches and models in psychopathology conceptualization research. *Nature Reviews Psychology*, *2*(10), Article 10. https://doi.org/10.1038/s44159-023-00218-4

Elsemüller, L., Schnuerch, M., Bürkner, P.-C., & Radev, S. T. (2023). *A deep learning method for comparing bayesian hierarchical models* (No. arXiv:2301.11873). arXiv. https://doi.org/10.48550/arXiv.2301.11873

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *eLife*, *10*, e65074. https://doi.org/10.7554/eLife.65074

Frazier, D. T., Robert, C. P., & Rousseau, J. (2020). Model Misspecification in Approximate Bayesian Computation: Consequences and Diagnostics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(2), 421–444. https://doi.org/10.1111/rssb.12356

Geng, H., Chen, J., Chuan-Peng, H., Jin, J., Chan, R. C. K., Li, Y., Hu, X., Zhang, R.-Y., & Zhang, L. (2022). Promoting computational psychiatry in China. *Nature Human Behaviour*, *6*(5), 615–617. https://doi.org/10.1038/s41562-022-01328-4

Ghaderi-Kangavari, A., Rad, J. A., & Nunez, M. D. (2023). A General Integrative Neurocognitive Modeling Framework to Jointly Describe EEG and Decision-making on Single Trials. *Computational Brain & Behavior*, *6*(3), 317–376. https://doi.org/10.1007/s42113-023-00167-4

Glickman, M., Moran, R., & Usher, M. (2022). Evidence integration and decision confidence are modulated by stimulus consistency. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-022-01318-6

Greenberg, D., Nonnenmacher, M., & Macke, J. (2019). Automatic posterior transformation for likelihood-free inference. *International Conference on Machine Learning*, 2404–2414.

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Habermann, D., Schmitt, M., Kühmichel, L., Bulling, A., Radev, S. T., & Bürkner, P.-C. (2024). *Amortized bayesian multilevel models* (No. arXiv:2408.13230). arXiv. https://doi.org/10.48550/arXiv.2408.13230

Hato, T., Schumacher, L., Radev, S. T., & Voss, A. (2024). *Lévy versus wiener: Assessing the effects of model misspecification on diffusion model parameters*. In Review. https://doi.org/10.21203/rs.3.rs-5276939/v1

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., & Louppe, G. (2022). *A trust crisis In simulation-based inference? Your posterior approximations can Be unfaithful* (No. arXiv:2110.06581). arXiv. https://doi.org/10.48550/arXiv.2110.06581

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11. https://doi.org/10.1016/j.copsyc.2018.10.008

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Kucina, T., Wells, L., Lewis, I., De Salas, K., Kohl, A., Palmer, M. A., Sauer, J. D., Matzke, D., Aidman, E., & Heathcote, A. (2023). Calibration of cognitive tests to address the reliability paradox for decision-conflict tasks. *Nature Communications*, *14*(1), 2234. https://doi.org/10.1038/s41467-023-37777-2

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

Miletić S., Boag R. J., Trutti A. C., Stevenson N., Forstmann B. U., & Heathcote A. (2021). A new model of decision processing in instrumental learning tasks. *eLife*, *10*, 1–55. https://doi.org/10.7554/elife.63055

Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *NeuroImage*, *197*, 93–108. https://doi.org/10.1016/j.neuroimage.2019.04.052

Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology*, *76*(1), 117–130. https://doi.org/10.1016/j.jmp.2016.03.003

O’Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*(1), 495–516. https://doi.org/10.1146/annurev-neuro-092019-100200

Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-72425-6

Papamakarios, G., & Murray, I. (2016). Fast ε-free inference of simulation models with bayesian conditional density estimation. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates, Inc.

https://proceedings.neurips.cc/paper_files/paper/2016/file/6aca97005c68f1206823815f66102863-Paper.pdf

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, *24*(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Radev, S. T., D'Alessandro, M., Mertens, U. K., Voss, A., Kothe, U., & Burkner, P.-C. (2021). Amortized bayesian model comparison with evidential deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. https://doi.org/10.1109/TNNLS.2021.3124052

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Kothe, U. (2022). BayesFlow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(4), 1452–1466. https://doi.org/10.1109/TNNLS.2020.3042395

Radev, S. T., Schmitt, M., Pratz, V., Picchini, U., Köthe, U., & Bürkner, P.-C. (2023). JANA: jointly amortized neural approximation of complex bayesian models. *Uncertainty in Artificial Intelligence*, 1695–1706.

Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., Köthe, U., & Bürkner, P.-C. (2023). BayesFlow: Amortized bayesian workflows with neural networks. *Journal of Open Source Software*, *8*(89), 5702. https://doi.org/10.21105/joss.05702

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction tune and IQ. *INTELLIGENCE*, *36*(1), 10–17. https://doi.org/10.1016/j.intell.2006.12.002

Rmus, M., Pan, T.-F., Xia, L., & Collins, A. G. E. (2024). Artificial neural networks for model identification and parameter estimation in computational cognitive models. *PLoS Computational Biology*, *20*(5), e1012119. https://doi.org/10.1371/journal.pcbi.1012119

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126. https://doi.org/10.1037/met0000275

Schumacher, L., Bürkner, P.-C., Voss, A., Köthe, U., & Radev, S. T. (2023). Neural superstatistics for bayesian estimation of dynamic cognitive models. *Scientific Reports*, *13*(1), 13778. https://doi.org/10.1038/s41598-023-40278-3

Schumacher, L., Schnuerch, M., Voss, A., & Radev, S. T. (2024). Validation and comparison of non-stationary cognitive models: A diffusion model application. *Computational Brain & Behavior*. https://doi.org/10.1007/s42113-024-00218-4

Straub, D., Niehues, T. F., Peters, J., & Rothkopf, C. A. (2025). *Inverse decision-making using neural amortized bayesian actors* (No. arXiv:2409.03710). arXiv. https://doi.org/10.48550/arXiv.2409.03710

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00079

van Geen, C., & Gerraty, R. T. (2021). Hierarchical bayesian models of reinforcement learning: Introduction and comparison to alternative methods. *Journal of Mathematical Psychology*, *105*, 102602. https://doi.org/10.1016/j.jmp.2021.102602

von Krause, M., Radev, S. T., & Voss, A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature Human Behaviour*, *6*(5), 700–708. https://doi.org/10.1038/s41562-021-01282-7

Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. https://doi.org/10.1016/j.jmp.2016.04.008

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, *111*(26), 9431–9436. https://doi.org/10.1073/pnas.1407756111

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wu, Y., Radev, S., & Tuerlinckx, F. (2024). *Testing and improving the robustness of amortized bayesian inference for cognitive models* (No. arXiv:2412.20586). arXiv. https://doi.org/10.48550/arXiv.2412.20586

Zammit-Mangion, A., Sainsbury-Dale, M., & Huser, R. (2024). Neural Methods for Amortized Inference. *Annual Review of Statistics and Its Application*. https://doi.org/10.1146/annurev-statistics-112723-034123

附录

1. 贝叶斯框架下的认知建模

1.1 贝叶斯认知模型的数学表示方式

在贝叶斯框架下，认知模型的模拟器或生成形式可以表示为:

$$D \sim p(D|\theta) \tag{1}$$

θ为模型参数，代表潜在认知状态。D为生成数据，代表在相应认知状态下所产生的行为或观测数据。$p(D|\theta)$为条件概率分布，表示在给定参数θ下观察到的数据D的概率，其暗含了认知加工的机制，即潜在认知状态对于加工过程和行为的影响。

从推论的角度来看，研究者关注的是如何从实验数据中推断潜在的心理过程，即进行参数估计。与传统的频率参数推断方法（如最小二乘法）相比，贝叶斯统计的吸引力在于它允许同时评估模型参数的估计值及其不确定性，而不仅仅是提供精确的点估计。参数的后验分布可以通过贝叶斯公式表示：

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{2}$$

$p(\theta|D)$为参数θ的后验分布，表示在获取数据D后对于参数θ的推断，其概率分布的形式蕴含了参数及其所代表的心理状态的不确定性。后验分布依赖于两个关键函数的准确评估：模型参数的先验分布$p(\theta)$，代表在未获得数据前对于参数分布的信念；以及似然函数$p(D|\theta)$，也可以写作$L(D|\theta)$[1]，代表了在已知参数的条件下数据出现的概率。然而，对于具有复杂认知机制的模型，受到生成模型逆问题所带来的困扰，确定其似然函数通常更加困难。最后，$p(D)$为贝叶斯证据(evidence)，也称为边际似然(marginal likelihood)，表示在所有参数范围内观察到数据D的概率的总和，数学上表示为$p(D) = \int p(D|\theta)p(\theta)d\theta$。

1.2 贝叶斯框架下的模型比较

在贝叶斯模型比较框架下，多个竞争生成模型被表示为集合$\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_j\}$，每个模型中参数的后验分布表示为，其中$J = 1{:}\,j$为模型序号：

---

[1] 可以发现，$p(D|\theta)$既可以表示生成模型，也可以表示似然函数。这是因为对于简单的概率分布而言，如正态分布，从分布中生成数据D，和反过来根据数据D计算似然值均相对容易实现。因此，为了避免概念上的混淆，我们将在后续的讨论中使用$L(D|\theta)$表示似然函数，而$p(d|\theta)$表示生成函数。注意，在一些统计和机器学习领域习惯将似然函数写作$L(\theta|D)$，而非$L(D|\theta)$。

$$p(\theta_j|D,\mathcal{M}_j)=\frac{p(D|\theta_j,\mathcal{M}_j)p(\theta|\mathcal{M}_j)}{p(D|\mathcal{M}_j)} \tag{3}$$

与公式2)相比，贝叶斯推断中的所有元素均受到模型的条件概率的影响。其中，边际似然(marginal likelihood)或贝叶斯证据(Bayesian evidence)为$p(D\mid\mathcal{M}_j)=\int p(D\mid\theta_j,\mathcal{M}_j)p(\theta_j\mid\mathcal{M}_j)\mathrm{d}\theta_j$，表示模型$\mathcal{M}_j$生成数据$D$的概率。边际似然作为模型选择的核心，为计算贝叶斯因子和模型后验概率奠定了基础。其中，贝叶斯因子通过边际似然的比率来量化模型间的相对支持度，定义为 $\mathrm{BF}_{jk}=\frac{p(x|\mathcal{M}_j)}{p(x|\mathcal{M}_k)}$，其中$\mathcal{M}_j,\mathcal{M}_k$分别代表两个竞争模型。当$\mathrm{BF}_{jk}>1$，表示在给定数据$D$的情况下，模型$\mathcal{M}_j$比模型$\mathcal{M}_k$更受青睐(胡传鹏等, 2018)。另一方面，结合边际似然和模型先验$p(\mathcal{M}_j)$可以计算得到模型后验概率，以直接量化在指定数据D下模型$\mathcal{M}_j$出现的概率，其数学形式为：

$$p(\mathcal{M}_j\mid D)=\frac{p(D\mid\mathcal{M}_j)p(\mathcal{M}_j)}{\sum_{j=1}^{J}p(D\mid\mathcal{M}_j)p(\mathcal{M}_j)} \tag{4}$$

贝叶斯因子和模型后验概率的计算均依赖于边际似然度，但由于边际似然度通常涉及高维积分，这导致其直接计算变得复杂或不可行。证据网络的核心思路在于利用神经网络直接从数据中估计模型后验概率，从而巧妙地规避了边际似然的计算难题。

2. 早期模拟推断推断方法数学表达形式

对于认知模型而言，即便模型具备明确的似然函数，模型参数的后验分布也可能不具备封闭形式的解。因此，研究者采用 MCMC 和变分推断(variantional inference, VI)等进行模型推断(Farrell & Lewandowsky, 2018; Forstmann & Turner, 2024)。但是当模型的似然函数不明确或难以计算时，研究者可能不得不放弃复杂的模型，或者采用避免计算似然函数的方法。近年来，研究者通过模拟器生成模拟数据来对真实观测数据进行推断，这种方法被称为无似然推断(likelihood-free inference)技术，也被称为基于模拟的推断(Cranmer et al., 2020)。当前的模拟推断有两种策略：近似贝叶斯计算和概率密度近似。

近似贝叶斯计算(Approximate Bayesian Computation，ABC)是针对贝叶斯模型设计的框架，已成为解决无似然推理问题的重要工具(Kangasrääsiö et al., 2019; Palestro et al., 2018; Turner & Zandt, 2018)。近似贝叶斯的核心思想在于通过生成和处理模拟数据，间接逼近难以直接计算的后验分布$p(\theta|D)$。其步骤包括：从先验分布中抽取参数样本；用这些参数生成模拟数据；计算模拟数据和实际数据的摘要统计量；比较这些统计量的差异；根据预设的阈值决定是否接受参数作为后验样本。重复这一过程，累积足够的后验样本，以此构建后验分布的近似(Palestro et al., 2018)。

部分研究者采用概率密度近似(Probablistic Density Approximation, PDA)以替代近似贝叶斯计算中的摘要统计，以实现更准确的参数估计(Holmes, 2015; Turner & Sederberg, 2014)。概率密度近似的核心思想在于利用整个模拟数据的分布信息来近似似然函数，称为伪似然(pseudo-likelihood)，记作$\hat{L}(\theta^*|\tilde{D})$。该方法需要先根据参数生成大量模拟数据，以覆盖该参数下似然分布，随后使用核密度估计方法来加权平均估计模拟数据的概率密度函数(Turner & Sederberg, 2014)。概率密度近似具有似然函数的优势，既适用于频率优化方法，也能与 MCMC 方法结合，用于可计算模型评估指标 (Holmes, 2015; Robert et al., 2011)。

2.1 近似贝叶斯计算

近似贝叶斯计算(Approximate Bayesian Computation，ABC)是一个算法系列，最基础的实现是基于拒绝的近似贝叶斯计算(rejection-based ABC)。其核心思想在于提供一种替代或绕过计算似然函数$L(\theta|D)$的方案，通过模拟数据的生成和处理，间接逼近难以直接计算的后验分布$p(\theta|D)$。该方法包括 5 个核心步骤(Palestro et al., 2018)：首先从先验分布$\pi(\theta)$中抽取参数θ^*的样本；进而利用参数生成模拟数据$\tilde{D} \sim p(D|\theta^*)$(见公式1))；同时计算模拟数据数据和真实数据的摘要统计量(summary statistic)，分别得到$S(\tilde{D})$和$S(D)$；随后比较模拟数据与观测数据的摘要统计之间距离(接近程度)，$d = \rho\left(S(\tilde{D}), S(D)\right)$[2]；最后，依据预设的距离度量$\epsilon$评估摘要统计距离$d$，如果$d < \epsilon$，则接受参数为后验样本，即$\theta_i = \theta^*$。重复整个流程直至累积到足够的后验样本量，进而构建出后验分布的近似。当$\epsilon \to 0$时，该方法推理的准确性会得到提高，但会使得接受概率降低，进而导致需要进行更多次数据模拟。反过来说，较大的ϵ值会提高获得后验样本效率，但代价是牺牲推理的质量(Cranmer et al., 2020)。

2.2 概率密度近似

鉴于摘要统计需要手动设置距离度量，部分研究者开始采用概率密度近似(Probablistic Density Approximation, PDA)以替代传统的摘要统计，旨在实现更为准确的参数估计(Holmes, 2015; Turner & Sederberg, 2014)。

概率密度近似是一种非参数估计方法，其核心思想在于运用整个数据的分布信息而非仅靠汇总统计量来形成对于似然函数的近似，称为伪似然(pseudo-likelihood)，记作$\hat{L}(\theta^*|\tilde{D})$。具体来说，该方法需要先根据参数生成一组模拟数据，$\tilde{D} = \{\tilde{d}_i\}$，其中$i$取 1 到

---

[2] 在近似贝叶斯计算方法中，计算摘要统计量距离的函数旨在衡量模拟数据摘要统计量与实际数据摘要统计量之间的差异，包括欧几里得距离、曼哈顿距离、切比雪夫距离、马氏距离、夹角余弦以及基于核的距离等(Bhattacharya et al., 2007)。

N，表示N个模拟数据。之后，利用该数据集构建概率密度函数，作为似然函数的近似，数学形式为：

$$L(\theta|d) \approx \hat{L}(\theta|d) \coloneqq \frac{1}{N}\sum_{i=1}^{N} K_h\left(d - \tilde{d_\iota}\right) \quad (5)$$

其中，d表示待估计似然值的真实数据，可以为单个数据点或多个数据；$K_h\left(d - \tilde{d_\iota}\right)$表示核密度估计(Kernel Density Estimation，KDE)方法。核密度估计是一种平滑技术，通过加权平均相邻数据点来估计未知观测数据的概率密度函数，这使得它能够基于现有数据集推断出新数据点的概率分布情况，从而用于密度估计(Turner & Sederberg, 2014)。

由于概率密度近似能近似似然函数的特性，使其展现出广泛的应用潜力。该方法不仅能直接用于频率优化方法，还能与 MCMC 方法结合。此外，概率密度近似所计算的近似似然值可进一步用于计算模型评估指标，使其在模型比较方面相对于近似贝叶斯计算方法具有一定的优势，这一议题已有相关讨论(Holmes, 2015; Robert et al., 2011)。

2.3 模拟推断方法与 MCMC 的异同

模拟推断通过模拟数据生成过程进行统计推断，适用于似然函数难以计算或不可解析表达的场景(Smith, 2018)。其核心思想是通过模拟数据与观测数据的匹配程度来推断参数后验分布。而 MCMC 通过构建马尔可夫链来逼近目标后验分布，依赖已知的似然函数和先验分布(Robert & Casella, 2004)。典型算法包括 Metropolis-Hastings、Gibbs 采样等。

模拟推断与 MCMC 在多个方面存在显著差异。首先，SBI 无需显式计算似然函数，而 MCMC 则需要。这使得 SBI 在处理复杂模型时更具灵活性。其次，SBI 的计算复杂度较高，需大量模拟数据，而 MCMC 的计算复杂度中等，依赖链的收敛速度。此外，SBI 的误差主要来源于模拟误差与近似阈值选择，而 MCMC 的误差则与链的混合速度和收敛性相关。在主要目标上，SBI 旨在近似后验分布，而 MCMC 则追求精确后验采样。

ABC 是 SBI 的核心技术，通过数据模拟与相似性度量绕过似然计算。Rejection-Based ABC 通过直接拒绝不满足阈值的样本进行推断，简单直观但效率低(Beaumont et al., 2002)。ABC-MCMC 结合了 MCMC 的马尔可夫链机制，提高了效率但可能陷入局部最优(Marjoram et al., 2003)。ABC-SMC 通过序贯迭代地收紧阈值，逐步逼近后验分布，适应复杂后验形态但计算成本高(Toni et al., 2009)。Kernel-Based ABC 使用核函数对距离进行加权，减少阈值敏感性但受核函数带宽选择影响(Blum & François, 2010)。

不同 ABC 方法在核心机制、阈值依赖、动态调整、计算效率及适用场景上各有特点。Rejection-Based ABC 适用于低维参数、快速验证场景；ABC-MCMC 适用于中等维度、局

部优化问题；ABC-SMC 适用于高维参数、多峰后验场景；Kernel-Based ABC 适用于平滑后验、减少阈值敏感性需求。

模拟推断方法与 MCMC 均旨在从数据中推断参数后验分布，但核心差异在于对似然函数的依赖程度。MCMC 适合模型可解析且似然易计算的场景；而 SBI/ABC 通过模拟绕过似然计算，适合复杂模型但需权衡计算成本与近似误差。在实际应用中，应根据模型特性、计算资源及精度要求选择合适的方法。

3. 神经网络相关核心技术

3.1 分摊(Amortization)

神经密度估计技术能够用于估计后验概率分布和似然函数，其特征在于分摊(Amortization)特性，其目标在于将计算资源集中用于训练一个单一模型，该模型能够一次性处理多个数据点，并且具有泛化到新数据的能力(见表 2)。分摊网络的优势在于一旦模型训练完成，可以对新的数据点进行快速推断，而无需进行额外训练(Radev et al., 2022, 2023)。

以神经后验估计为例，分摊策略包括三个主要步骤：在模拟阶段，生成大量训练数据对$\{\tilde{\theta}_i, \tilde{D}_i^{1:M}\}_{i=1}^{N}$，其中$i$取 1 到$N$，表示$N$份数据对，$\tilde{D}_i$由 1 到$M$个数据点构成；在训练阶段，调整神经网络权重$\phi$使得神经网络$q_\phi$学习数据$D$和概率密度$p(\theta|D)$的映射；在推理阶段，输入真实观测数据$D$，运行神经网络后得到参数后验样本$p(\theta|D) := q_\phi(\theta|D)$。分摊策略的关键在于训练和推理完全分离，一旦网络训练完成，神经网络可以离线运行，以极快的速度获得推理，即训练过程分摊了推理的高额运算(Radev et al., 2020)。此外，由于模型是基于整个先验预测分布产生的数据集上训练的，因此它通常具有更好的泛化能力，能够更好地处理新数据(Radev et al., 2020)。

3.2 标准化流(Normalizing Flow, NF)

标准化流(Normalizing Flow, NF, Papamakarios, 2019; Papamakarios et al., 2021)采用生成模型架构，可实现对条件密度的直接估计，并且其所需的训练数据更少。

标准化流作为一种可逆生成模型，执行无监督学习任务，无需借助核密度估计计算似然值作为训练标签。其核心思想在于通过连续的、可逆的变换将一个简单的分布$z \sim \pi(z)$(如高斯分布)转化为更复杂的数据分布。

在估计似然函数时，标准化流通过条件可逆神经网络(conditional invertible neural network, cINN)实现对条件概率密度的估计(Durkan et al., 2019; Papamakarios et al., 2021)，其数学形式为：

$$L_{\phi}(D \mid \theta) = \pi(z_D)\left|\det\left(\frac{\partial f^{-1}}{\partial D}\right)\right| \quad (6)$$

标准化流中，似然函数$L_{\phi}(D \mid \theta)$由神经网络参数化，其权重为ϕ，通过一个简单分布$\pi(z_D)$变换得到。其中，f^{-1}为可逆变换，可以将数据映射到简单分布的潜在空间，即$z_D = f(D)$，反之亦然，$D = f^{-1}(z_D)$。$\left|\det\left(\frac{\partial f^{-1}}{\partial D}\right)\right|$为雅可比矩阵的行列式，为可逆变换的局部缩放因子，衡量数据D到z_D映射过程的缩放性。正如其名，“标准化流”由一系列的可逆神经网络模块串接形成，其作用是首先将数据标准化为潜变量，然后通过一系列变换(即“流”)将其转化为目标分布(Papamakarios et al., 2021)。因此，标准化流的关键在于定义和学习可逆变换。

在估计参数后验分布$p(\theta \mid D)$时，可以表示为在给定数据D的条件下参数θ和潜在变量z_{θ}(来自于简单分布)之间的标准化流：

$$p_{\phi}\left(\theta \,\middle|\, \mathcal{H}_{\eta}(D)\right) = p(z_{\theta})\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right| \quad (7)$$

其中，f^{-1}为可逆变换；雅各比矩阵的行列式$\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right|$衡量了简单分布到后验分布的缩放性。

标准化流已发展出多种高效的可逆变换方法，包括仿射(affine)，多项式样条(polynomial spline)变化，和立方样条(cubic spline)变换(Durkan et al., 2019; Papamakarios et al., 2021)。MNLE 采用了神经样条流(neural spline flow, NSF)的架构，通过几个单调有理二次样条(monotonic rational-quadratic splines)模块实现对似然密度的估计。例如，对于同时包括反应时和反应，该模型可以采用包含两个网络的混合神经网络架构，其中一个网络负责离散变量，针对行为决策反应等离散数据，另一个网络负责连续变量，针对反应时等连续测量数据。

3.3 LANs 训练特点

Fengler 等人(2021)提出“似然近似网络”(Likelihood Approximate Networks, LANs)，以快速和准确地估计认知模型的似然函数。LANs 采用多层感知机(Multilayer Perceptrons, MLPs)来学习参数和数据到对数似然函数的映射 $f_{\mathrm{MLP}(\phi)}: (\theta, D) \to \log L\,(D|\theta)$，其中$f_{\mathrm{MLP}(\phi)}$代表具有权重参数$\phi$的多层感知机神经网络。LANs 的训练过程包含两个关键步骤：数据准备和训练。

在数据准备阶段，首先采集一组参数 $\tilde{\theta}_i$；接着根据参数生成模拟数据$\tilde{D}_i^{1:M}$；随后基于核密度估计方法(见概率密度近似小节)计算每个生成数据点的对数似然函数值

$\log L\left(\widetilde{D}\middle|\tilde{\theta}\right)_i^{1:M}$；重复以上步骤 N 次得到训练数据集$\left\{\tilde{\theta}_i, \widetilde{D}_i^{1:M}, \log L\left(\widetilde{D}\middle|\tilde{\theta}\right)_i^{1:M}\right\}_{i=1}^N$。为确保网络能够有效学习数据的多样化概率分布，训练数据集中的参数从匀分布中抽取，包含 150 万个参数样本，每个样本再输入模拟器生成 10 万个数据点(Fengler et al., 2021)。因此，训练集共包含 1500 亿个带标签的样本以供网络训练。

在模型训练阶段，LANs 执行监督学习任务，其中输入为参数$\tilde{\theta}$和数据$\widetilde{D}$，输出为对数似然函数的估计$\log \tilde{L}\left(\widetilde{D}\middle|\tilde{\theta}\right)$。通过优化网络权重$\phi$以最小化对数似然估计$\log \tilde{L}\left(\widetilde{D}\middle|\tilde{\theta}\right)$与实际对数似然值$\log L\left(\widetilde{D}\middle|\tilde{\theta}\right)_i^{1:M}$之间的差异，通过 Huber 损失函数衡量。训练通过随机梯度下降的反向传播算法进行，并采用 Adam 优化算法来调整网络权重(Kingma & Ba, 2017)。

3.4 MNLE 训练特点

为了克服 LANs 的局限性，Boelts 等(2022)提出了混合神经似然估计(Mixed Neural Likelihood Estimation, MNLE)。MNLE 采纳标准化流架构，可实现对条件密度的直接估计。相较于 LANs 使用类似“回归”的方式建立数据和参数到似然值的映射，MNLE 可以直接对于似然函数进行估计，并且其所需的训练数据更少(Boelts et al., 2022)。

MNLE 同样涉及两个训练阶段。在数据准备阶段，MNLE 需要构造数据对集$\left\{\tilde{\theta}_i, \widetilde{D}_i\right\}_{i=1}^N$。相较于 LANs，MNLE 具有两点优势：首先，MNLE 不再依赖 KDE 生成似然值作为标签；其次，MNLE 中的参数和数据点之间是一一对应的关系，而无需根据每组参数生成大量数据点$\widetilde{D}_i^{1:M}$。在训练阶段，MNLE 执行无监督学习任务，通常采用 Adam 优化器和最大似然损失函数(maximum likelihood loss)，即$\sum_{i=1}^N \log L_\phi\left(D_i|\theta_i\right)$。

MNLE 的优势在于其模拟效率。相较于 LANs 需要百亿级别的模拟数据，MNLE 仅需万级别的数据量即可达到或超越 LANs 的准确性。此外，作为一种生成模型，MNLE 不仅能够估计似然值，还能够充当模拟器，可以迅速生成合成数据，从而避免了实际模拟运算的需求。这一特性在评估模型性能和计算模拟器内部信息(如参数的梯度信息)方面显示出潜在优势。更重要的是，MNLE 已被整合至 Python 开源工具包“sbi”中，这为在认知神经科学领域中实现无似然推断方法提供了基础(Tejero-Cantero et al., 2020)。

3.5 Bayesflow

Radev 等人(2020)首次将神经后验估计方法应用于认知建模领域，提出了 BayesFlow 框架，该框架结合了标准化流和分摊策略，旨在通过一次性训练实现对认知模型参数的快速和高效推断(Radev et al., 2020, 2022)。

与 MNLE 的不同之处在于，BayesFlow 引入了摘要统计网络(summary statistics network)，即$\mathcal{H}_\eta$，以解决基于深度学习的近似贝叶斯计算不允许试次数量变化的问题，该

网络可以将任意长度的数据$D^{1:M}$(假设这些数据来自于独立同分布，并且具有置换不变性)映射为固定长度的嵌入向量$\mathcal{H}_\eta(D)^{1:K}$，其中$K$代表摘要统计维度的潜在维度。因此，参数后验的标准化流并不是直接基于数据进行变换，而是基于数据的摘要统计和参数先验进行可逆变换，即$z_D = f(\theta; \mathcal{H}_\eta(D))$和$\theta = f^{-1}(z_D; \mathcal{H}_\eta(D))$ (Radev et al., 2022)。

Bayesflow 执行无监督学习任务，所需的训练数据集形式为$\{\tilde{\theta}_i, \widetilde{D}_i^{1:M_i}\}_{i=1}^{N}$，其中$\widetilde{D}_i^{1:M_i}$用于生成摘要统计$\mathcal{H}_\eta(D)^{1:K}$。需要注意的是，每对数据的样本数量$M_i$存在差异，这使得训练后的模型能够适应各种真实样本数量。在训练期间，$\{\tilde{\theta}_i, \mathcal{H}_\eta(D)^{1:K}\}$共同输入条件可逆神经网络生成$z_D$，之后通过最小化 KL(Kullback-Leibler)散度，即$\underset{\phi,\eta}{min}\mathbb{KL}(p(\theta|D)||p_\phi(\theta|\mathcal{H}_\eta(D)))$，共同优化摘要统计和条件可逆神经网络两个网络，实现对于真实后验的近似(Radev et al., 2022)。

3.6 证据网络训练特点

Radev 等人(2021)首次在认知建模领域中使用神经模拟比较方法，聚焦于解决贝叶斯模型比较(Bayesian model comparision，BMC)问题。他们提出证据网络(evidential network)，借鉴神经似然比估计的思想，可以直接计算近似贝叶斯因子(Bayes factor, BF)或模型后验概率(posterior model probabilities, PMP)，并可扩展至对于多个模型的比较。该方法完全基于模拟，具有分摊性质，并且无需为各模型训练独立神经网络。

证据网络训练的关键在于准备训练数据$\{\mathcal{M}_j, \widetilde{D}_i^{1:M_i}\}_{i=1}^{N}$：首先根据模型先验$p(\mathcal{M}_j)$采样得到模型$\mathcal{M}_j$；随后根据模型对应的参数先验分布$p(\theta_i|\mathcal{M}_j)$采样得到参数$\theta_i$；最后再根据参数生成数据$\widetilde{D}_i^{1:M_i}$，其中数据的样本量为$1{:}M$。证据网络执行监督学习分类任务，证据网络接受输入数据$\widetilde{D}_i^{1:M_i}$，并输出模型的后验概率$p(\mathcal{M}_j \mid D)$，可表示为$f_\phi: D \to p(\mathcal{M}_j \mid D)$，其中输出为$J$维向量，对应每个模型出现的概率。值得注意的是，为了量化神经网络估计的不确定性，证据网络不是直接输出，而是作为狄利克雷分布的浓度参数，结合模型先验来形成二阶概率分布。因此，证据网络的输出可被视为绝对证据，而狄利克雷分布的二阶概率分布可用于定量描述神经网络估计的不确定性(Radev et al., 2021)。

4. 神经模拟推断的三种类型

神经模拟推断包括三种应用类型(见表 I)，神经似然估计、神经后验估计，以及基于神经网络的模型比较。正文主要阐述了神经后验估计的技术细节及其原理。接下来，文章将依次介绍神经似然估计方法，以及如何利用神经网络进行模型比较。

表 I 神经模拟推断的应用比较

	神经似然估计	神经后验估计	神经模型比较
描述	根据输入的先验参数和给定的观测数据输出似然值 $L_\phi(D \mid \theta)$	根据给定观测数据输出参数后验分本的样本 $p_\phi(\theta \mid D)$	根据输入的数据给出对应的模型证据 $p_\phi(\mathcal{M}_j \mid D)$
优点	可结合 MCMC 等算法进行参数估计； 并且可以结合后验参数和桥采样等算法用于计算模型比较评估指标； 可用于分层模型	无需进行 MCMC 采样，能快速获得参数后验分布样本	通过单个神经网络模型考虑多个认知模型的比较； 无需重新训练，可通过调试神经网络扩展到新的认知模型
缺点	受到优化算法(例如 MCMC)的限制； 模型推理速度较慢	无法获得似然值，无法直接进行模型评估和比较； 需根实验设计训练专门的神经网络	无法对模型参数进行推断
应用示例	(LANs, Fengler et al., 2021) (MNLE, Boelts et al., 2022) (JANA, Radev, Schmitt, Pratz, et al., 2023)	(Bayesflow, Radev, Voss, et al., 2020) (JANA, Radev, Schmitt, Pratz, et al., 2023) (ANN, Rmus et al., 2024)	(DEN Radev et al., 2021) (DHMC, Elsemüller et al., 2023) (ANN, Rmus et al., 2024)
工具	(LANs, Fengler et al., 2021) (sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)	(sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)	(sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)

4.1 神经似然估计

神经似然估计方法被引入认知建模中，作为概率密度近似的替代模型，其目标在于免去每次计算似然值时需重复生成模拟数据的过程，同时所计算的似然值可用于频率推断或者 MCMC 等方法。Fengler 等人(2021)提出“似然近似网络”(Likelihood Approximate Networks, LANs)，以快速和准确地估计认知模型的似然函数。

LANs 采用多层感知机(Multilayer Perceptrons, MLPs)来学习参数和数据到对数似然函数的映射 $f_{\mathrm{MLP}(\phi)}: (\theta, D) \to \log L\,(D|\theta)$，其中$f_{\mathrm{MLP}(\phi)}$代表具有权重参数$\phi$的多层感知机神经网络。LANs 执行监督学习任务，其中输入为参数$\tilde{\theta}$和数据$\tilde{D}$，输出为对数似然函数值 $\log \tilde{L}\,(\tilde{D}|\tilde{\theta})$。通过优化网络权重$\phi$以最小化对数似然估计$\log \tilde{L}\,(\tilde{D}|\tilde{\theta})$与实际对数似然值（标签数据）$\log L\,(\tilde{D}|\tilde{\theta})_i^{1:M}$之间的差异，最终习得数据到对数似然函数的映射。

训练后的 LANs 可以替代似然函数，与现有的 MCMC 法结合使用 (Fengler et al., 2021)。LANs 的显著优势在于能够评估单个观测值的点似然性(pointwise likelihood)，这使其适用于分层模型和分层线性模型。尽管 LANs 能显著提升对于复杂认知模型推断的能力，但仍然存在局限性。首先，LANs 的训练成本高昂，需要数十亿次模拟(Fengler et al., 2021)，这限制了其广泛应用，因为大多数用户无法实施如此大规模的训练。此外，LANs 并非直接进行密度估计，而是采用类似于回归模型的方法建立数据和参数到似然值的映射。这种模型架构一方面要求大量参数以覆盖参数空间，另一方面，即便能够覆盖，也难以在参数空间中实现平滑的估计(Boelts et al., 2022)。

为了克服 LANs 的局限性，Boelts 等(2022)提出了混合神经似然估计(Mixed Neural Likelihood Estimation, MNLE)。MNLE 采用标准化流(Normalizing Flow, Papamakarios, 2019; Papamakarios et al., 2021)，可实现对条件密度的直接估计。不同于 LANs 使用类似“回归”的方式建立数据和参数到似然值的映射，MNLE 直接估计似然函数，且其所需的训练数据更少(Boelts et al., 2022)。与 LANs 相比，MNLE 具有两点优势：首先，MNLE 采用生成模型构建执行无监督学习任务，因此 MNLE 不再依赖似然值作为标签；其次，MNLE 中的参数和数据点之间是一一对应的关系，而无需根据每组参数生成大量数据点，因此其效应更高。Boelts 等(2022)的研究结果显示，MNLE 仅需万级别的数据量即可达到或超越 LANs 基于百亿模拟数据所达到的准确性。MNLE 已被整合至 Python 开源工具包“sbi”中，这为在认知神经科学领域中实现无似然推断方法提供了基础(Tejero-Cantero et al., 2020)。

LANs 和 MNLE 表明神经似然估计在认知建模领域的可行性。然而，这类方法仍有改进的空间。首先，神经似然估计在实际运用时仍然需要使用优化或近似方法，如 MCMC，这导致其推断效率仍然不高。其次，如果神经网络没有经过充分的训练，网络本身的噪声加上近似方法的噪声会进一步放大估计的不准确性。此外，尽管神经似然估计通常能利用图形处理单元(GPU)加速推断的过程，但当需要并行运行多条 MCMC 链时，这可能会与 GPU 的架构不匹配，从而降低其计算效率。

神经似然估计对参数的推断是间接的，需要与 MCMC 等方法相结合，而神经后验估计则专注于直接对参数的后验分布进行估计。神经后验估计可以被视为基于深度学习的近似贝叶斯计算的改进，旨在解决后者面临的两个核心挑战：一是推断过程中倾向于估计后验分布的摘要统计量而非完整的后验密度；二是在处理数据时，仅能针对预设数量的观测值进行参数估计，缺乏对数据量的适应性。

4.2 神经模型比较

神经后验估计类方法的最大局限在于缺乏灵活性，每当参数有变化时均需要重新训练神经网络。例如，当需要根据实验条件比较参数时，意味着为额外的条件增加参数，而模型估计的参数数量是固定的，因此，增加实验条件意外着需要重新训练一个类似的模型。同样，神经后验估计无法完成具有约束参数的推断需要，如无法根据先前研究的最优参数约束当前模型，导致无法评估数据对特定参数的影响程度。此外，目前的神经后验估计亦无法处理分层模型，仅能针对单个被试或条件进行估计。相较之下，神经似然估计方法在这方面更具优势，无需针对不同先验、实验条件和分层模型进行额外训练。最后，神经后验估计通常依赖于摘要统计网络，这对推断性能至关重要，但在某些实际场景中，构建合适的汇总网络可能存在困难(Radev et al., 2022)。

模型评估和比较是认知建模中另一个重要的环节（见郭鸣谦等，2024），但神经似然与神经后验估计均不适用于模型评估和比较，原因在于神经后验估计跳过计算似然值和边际似然，而神经似然估计虽然能够用于计算似然值和边际似然，但处理复杂模型时计算成本高昂(Mancini et al., 2023; Radev et al., 2021)。基于神经似然估计的模型比较的另一个局限在于，需要为每个待比较的模型单独训练神经网络，增加了实现的成本。

针对这些问题，神经模型比较的思路被提出，旨在基于模拟方法识别给定数据最有可能来自于哪个生成模型，从而避免进行参数估计或计算似然和边际似然(Radev et al., 2021)。

神经模型比较方法的核心在于训练一个分类器，用以执行监督学习任务，其中输入为多个生成模型产生的模拟数据，输出则是与这些模型相对应的标签。例如，Radev 等人(2021)提出证据网络(Evidential Network)，该网络接受输入数据$\widetilde{D}_i^{1:M_i}$，并输出模型$\mathcal{M}_j$的后验概率$p\left(\mathcal{M}_j \mid D\right)$，可表示为$f_\phi: D \rightarrow p\left(\mathcal{M}_j \mid D\right)$，其$J$为模型数量。这意味着一旦模型完成训练，就能够根据输入的真实数据，推断出最有可能产生数据的模型。

证据网络的优势在于能够近似贝叶斯因子和模型后验概率，以此估计绝对和相对证据并量化估计的不确定性(Radev et al., 2021)。此外，证据网络采用分摊方法避免了对每个候选模型和数据集单独进行训练的步骤，且无需更改网络结构，仅需调整输出层即可实现对于新模型的评估(Radev et al., 2021)。

证据网络的局限在于，其基于边际似然进行模型比较，因而无法依据后验预测进行模型的评估和比较（见郭鸣谦等，2024）。此外，证据网络不适用于处理具有高维参数空间的模型，尤其是对于贝叶斯分层模型而言。针对于此，Elsemüller 等(2023)提出深度分层模型比较(deep hierarchical model comparision, DHMC)方法，旨在比较具有多层结构和高维参数

的贝叶斯模型。DHMC 方法在证据网络的基础上使用两组不变性网络分别处理试次和被试数量不一致的问题，在多种贝叶斯模型比较场景中展现出与桥接抽样技术相当或更优的性能，适用于广泛的模型比较任务。

参考文献

胡传鹏, 孔祥祯, Eric-Jan Wagenmakers, Alexander Ly, 彭凯平. (2018). 贝叶斯因子及其在 JASP 中的实现. *心理科学进展*, *26*(06), 951–965. https://doi.org/10.3724/SP.J.1042.2018.00951

Bhattacharya, A., Sarkar, B., & Mukherjee, S. K. (2007). Distance-based consensus method for ABC analysis. *International Journal of Production Research*, *45*(15), 3405–3420. https://doi.org/10.1080/00207540600847145

Boelts, J., Lueckmann, J.-M., Gao, R., & Macke, J. H. (2022). Flexible and efficient simulation-based inference for models of decision-making. *eLife*, *11*, e77220. https://doi.org/10.7554/eLife.77220

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *eLife*, *10*, e65074. https://doi.org/10.7554/eLife.65074

Holmes, W. R. (2015). A practical guide to the probability density approximation (PDA) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68–69*, 13–24. https://doi.org/10.1016/j.jmp.2015.08.006

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization* (No. arXiv:1412.6980). arXiv. https://doi.org/10.48550/arXiv.1412.6980

Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-72425-6

Papamakarios, G. (2019). *Neural density estimation and likelihood-free inference* (No. arXiv:1910.13233). arXiv. http://arxiv.org/abs/1910.13233

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research, 22*(57), 1–64.

Radev, S. T., D'Alessandro, M., Mertens, U. K., Voss, A., Kothe, U., & Burkner, P.-C. (2021). Amortized bayesian model comparison with evidential deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. https://doi.org/10.1109/TNNLS.2021.3124052

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Kothe, U. (2022). BayesFlow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning Systems, 33*(4), 1452–1466. https://doi.org/10.1109/TNNLS.2020.3042395

Radev, S. T., Schmitt, M., Pratz, V., Picchini, U., Köthe, U., & Bürkner, P.-C. (2023). JANA: jointly amortized neural approximation of complex bayesian models. *Uncertainty in Artificial Intelligence*, 1695–1706.

Radev, S. T., Voss, A., Wieschen, E. M., & Bürkner, P.-C. (2020). *Amortized bayesian inference for models of cognition* (No. arXiv:2005.03899). arXiv. https://doi.org/10.48550/arXiv.2005.03899

Robert, C. P., Cornuet, J.-M., Marin, J.-M., & Pillai, N. S. (2011). Lack of confidence in approximate bayesian computation model choice. *Proceedings of the National Academy of Sciences*, *108*(37), 15112–15117. https://doi.org/10.1073/pnas.1102900108

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., & Macke, J. H. (2020). *SBI -- a toolkit for simulation-based inference* (No. arXiv:2007.09114). arXiv. https://doi.org/10.48550/arXiv.2007.09114

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review, 21*(2), 227–250. https://doi.org/10.3758/s13423-013-0530-0

Neural simulation-based inference: a neural network and simulation-based inference approach to cognitive modelling

Wanke Pan[1,2], Hu Chuan-Peng[1,2]

[1] School of Psychology, Nanjing Normal University, Jiangsu, Nanjing 210097, China

[2] Adolescent Education and Intelligence Support Lab of Nanjing Normal University, Laboratory of Philosophy and Social Sciences at Universities in Jiangsu Province

Abstract

Cognitive computational modeling quantifies human mental processes using mathematical frameworks, translating cognitive theories into testable hypotheses. Modern cognitive modeling involves four interconnected stages: defining models by formalizing symbolic theories into generative computational frameworks, collecting data through hypothesis-driven experiments, inferring parameters to quantify cognitive processes, and evaluating or comparing models. Parameter inference, a critical step bridging models and data, traditionally relies on maximum likelihood estimation (MLE) and Bayesian methods like Markov Chain Monte Carlo (MCMC). These approaches depend on explicit likelihood functions, which become computationally intractable for complex models—such as those with nonlinear parameters (e.g., learning dynamics) or hierarchical/multimodal data structures.

To address these challenges, simulation-based inference (SBI) emerged, leveraging parameter-data mappings via simulations to bypass likelihood calculations. Early SBI methods, however, faced computational redundancy and scalability limitations. Recent advances in neural simulation-based inference (NSBI), or neural amortized inference (NAI), harness neural networks to pretrain parameter-data relationships, enabling rapid posterior estimation.

Despite its advantages, NSBI remains underutilized in psychology due to technical complexity. This work focuses on neural posterior estimation, one of three NSBI approaches alongside neural likelihood estimation and neural model comparison. Neural posterior estimation operates in two phases: training and inference. During training, parameters are sampled from prior distributions, and synthetic data are generated using the model; a neural network then learns to approximate the true posterior from these training pairs. In the inference stage, real data are input to the trained network to generate parameter samples. The BayesFlow framework enhances neural posterior estimation by integrating normalizing flows—flexible density estimators—and summary

statistic networks, enabling variable-length data handling and unsupervised posterior approximation. Its GPU-accelerated implementation further boosts efficiency.

Neural posterior estimation has expanded the scope of evidence accumulation models (EAMs), one of the most widely used framework in cognitive modeling. First, it enables large-scale behavioral analyses, as demonstrated by von Krause et al. (2022), who applied neural posterior estimation to drift-diffusion models (DDMs) for 1.2 million implicit association test participants. By modeling condition-dependent drift rates and decision thresholds, they revealed age-related nonlinear cognitive speed changes, peaking at age 30 and declining post-60. Neural posterior estimation completed inference in 24 hours versus MCMC's 50+ hours for a small subset, demonstrating its scalability.

Second, neural posterior estimation supports dynamic decision-making frameworks, exemplified by Schumacher et al. (2023), who combined high-level dynamics with low-level mechanisms using recurrent neural networks (RNNs). Their simultaneous estimation of hierarchical parameters achieved over 0.9 recovery correlations and superior predictive accuracy compared to static models.

Finally, neural posterior estimation facilitates neurocognitive integration, as shown by Ghaderi-Kangavari et al. (2023), who linked single-trial EEG components (e.g., CPP slope) to behavior via shared latent variables like drift rate. This approach circumvented intractable likelihoods and revealed associations between CPP slope and non-decision time.

NSBI enhances cognitive modeling by enabling efficient analysis of complex, high-dimensional datasets. Its key limitations include model validity risks (biased estimates from incorrect generative assumptions), overfitting concerns (overconfident posteriors on novel data), and upfront training costs for amortized methods. Future work should refine validity checks—such as detecting model misspecification—and develop hybrid inference techniques. NSBI's potential extends to computational psychiatry and educational psychology, promising deeper insights into cognition across domains. By addressing complexity barriers, NSBI could democratize advanced modeling for interdisciplinary research, advancing our understanding of human cognition through scalable, data-driven frameworks.

Keywords: Cognitive Modeling, Generative Models, Bayesian, Simulation-based Inference, Neural Network, Neural Amortized Bayesian Inference